Joint reanalysis of 29 correlated SNPs supports the role of PCLO/Piccolo as a causal risk factor for major depressive disorder

Bochdanovits, Z.; Verhage, M.; Smit, A.B.; de Geus, E.J.C.; Posthuma, D.; Boomsma, D.I.; Penninx, B.W.J.H.; Hoogendijk, W.J.G.; Heutink, P.

published in
Molecular Psychiatry
2009

DOI (link to publisher)
10.1038/mp.2009.37

Molecular Psychiatry (2009) 14, 650–652

www.nature.com/mp

LETTER TO THE EDITOR

Joint reanalysis of 29 correlated SNPs supports the role of PCLO/*Piccolo* as a causal risk factor for major depressive disorder

Molecular Psychiatry (2009) **14,** 650–652;
doi:10.1038/mp.2009.37

The first genome-wide association study (GWAS) for major depressive disorder (MDD) has implicated the pre-synaptic protein, *Piccolo*, but results from multiple replication cohorts remained inconclusive. Here, we present a reanalysis of the *Piccolo* replication study and argue that the data favor a non-synonymous coding single nucleotide polymorphism (SNP) in a well-characterized protein domain to be a causal risk factor for major depression.

In the first GWAS for MDD, multiple SNPs within the PCLO gene that codes for the pre-synaptic protein, *Piccolo*, reached nominal *P*-values in the magnitude of 10^{-7}.[1] One of the most strongly associated SNPs was rs2522833, a non-synonymous coding SNP in a well-characterized protein domain, the C2-domain, of *Piccolo*. A panel of 30 SNPs was followed up in a finemapping/replication effort that involved five additional MDD cohorts. In the replication stage, a nominally significant *P*-value was observed for rs2522833 in one cohort. Although these results are highly suggestive, the original finding did not reach genome-wide significance, and after permutation-based correction for testing multiple correlated SNPs and cohorts, the replication study was not conclusive.[1]

Here, we present a reanalysis of the PCLO replication study and conclude that Sullivan *et al*. provide convincing evidence to claim the association of rs2522833, a potentially causal variant in a biologically plausible candidate gene, with MDD. Our approach is based on two observations. First, finemapping studies focus on a limited number of correlated SNPs. All tested SNPs are expected to reflect the true association of the unknown causal variant proportional to their LD with it. This phenomenon is better known as the 'Fundamental Theorem of the HapMap',[2] and although it has received some criticism,[2,3] it is generally accepted that the observed effect at a neutral SNP is the product of the causal effect and the correlation between the causal and neutral loci (that is, LD). Second, given such correlated SNP data, it has been suggested before that a joint analysis of all markers together is most powerful for detecting a true association. A joint analysis *de facto* summarizes the information from many neighboring SNPs to provide one joint test of association (for review, see Bacanu[4]), hence no correction for testing multiple SNPs is needed.

A closer examination of the results reported by Sullivan *et al*. shows that the data indeed concur with the 'Theorem of the HapMap'. Sullivan *et al*. report in Table 4a the *Z*-scores of the association between each individual SNP and MDD in the original cohort used for the GWAS. From the HapMap data, the LD is known between rs2522833 and 28 of the remaining 29 SNPs. Interestingly, the correspondence between the absolute value of the *Z*-scores observed by Sullivan *et al*. and the square root of r^2 from HapMap is striking (Figure 1). Most of the observed association between the genotype data and MDD is explained ($r = 0.92$) by the association of rs2522833 with MDD and the local LD structure of *PCLO*. The data are consistent with the hypothesis that either rs2522833 or a variant in very high LD with rs2522833 is a causal risk factor for MDD. It should be noted that an earlier study has confirmed that the local LD structure in the PCLO region is highly similar between the HapMap CEU panel and the Dutch population sample used in this study.[5]

On the basis of the above, we propose that the replication data can be analyzed by a joint measure of association and local LD, similar to a recently reported method.[6] The specific hypothesis to address is the following: the non-synonymous coding SNP, rs2522833, is causally related to MDD, and the neighboring SNPs show associations proportional to their LD with rs2522833. To make use of the summary statistics reported in Table 4a by Sullivan *et al*., and the representation of the 'Theorem of the HapMap' depicted in Figure 1, we estimated the joint probability distribution of the *Z*-score for rs2522833 (Z) and the correlation coefficient between *Z*-scores and LD relative to rs2522833 (C) under the null hypothesis of no association of rs2522833 with MDD. This joint distribution was determined empirically from simulations using GWAsimulator.[7] From this empirical distribution, the joint probability was calculated for the observed values of $Z = T_2$ and $C = T_1$.

$$P(\mathrm{correl}(\mathrm{abs}\{Z_{1..n}\}, \mathrm{abs}\{r_{1..n,x}\}) > T_1, Z_x > T_2) \quad (1)$$

This probability is not the *P*-value for the test, because combinations other than the observed values, $Z = T_2$ and $C = T_1$, can also reach joint probabilities equally low or lower. The correct *P*-value is given by the proportion of all possible combinations of Z and C

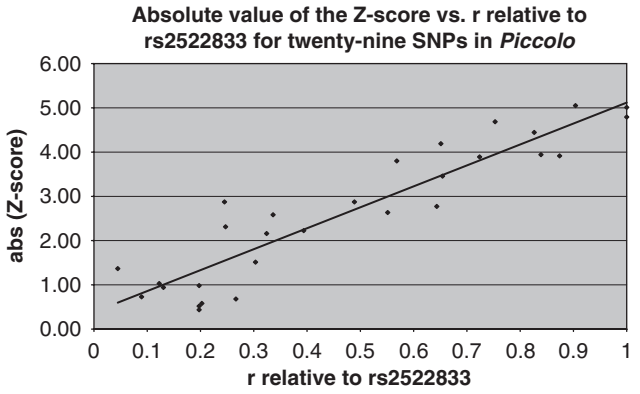

Figure 1 Linear fit between the observed genetic effects on MDD at the markers and the absolute value of the correlation between markers and the potentially causal variant, rs2522833. Although not the *Z*-score but the correlation coefficient between marker and phenotype is expected to scale linearly with LD, for correlation coefficients up to $r \sim 0.7$ (an unrealistically high genotype–phenotype correlation for a complex trait), *Z*-scores are a good proxy for r (a simulation in Visual Basic Macro in Excel is available from the first author to show this). Note that using absolute values implicitly assumes (in agreement with the theory) that the direction of the association with the phenotype aligns with the direction of LD between single nucleotide polymorphism (SNPs). Assuming this allows for the least manipulation of earlier published data, and because the simulation study presented below incorporates this limitation, our conclusions will not be biased.

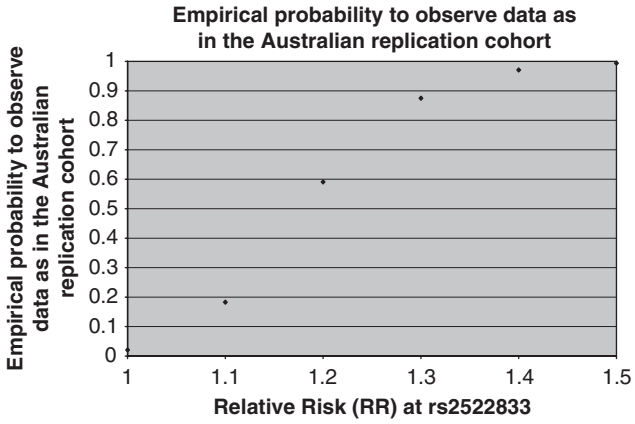

Figure 2 Empirical probability to jointly observe a *Z* and a *C*, which is equally or less likely than the observed values, T_1 and T_2, in the Australian replication cohort. Simulations were carried out using GWAsimulator conditioning on the sample size of the replication cohort and the local LD structure between the 29 single nucleotide polymorphism (SNPs) in the study. Realistic genetic effect sizes were modeled assuming multiplicative penetrance. The null hypothesis of no replication is represented by relative risk (RR) = 1. In all, 10 000 replicates were simulated, except for RR = 1 in which 50 000 replicates were used to more accurately estimate the low empirical probability under the null hypothesis.

with a joint probability lower than that which is observed for $Z = T_2$ and $C = T_1$.

Sullivan *et al*. report on five replication cohorts. One of these five was *a posteriori* described to be more similar to the original sample than the others because both studies included population-based cases, and controls were selected for low liability for MDD. In this (Australian) cohort, a nominally significant replication of the association between MDD and rs2522833 was observed: $Z = 2.19$ (T_2). Our parameter T_1 is 0.61 for this cohort. The joint probability is 0.008 and the *P*-value to observe such unlikely combinations of Z and C is 0.02 (Figure 2). Although the other four cohorts did not show evidence of association (data not shown), the present reanalysis of data favors the alternative hypothesis that rs2522833 (or an unknown SNP in strong LD with it) is a causal risk factor for MDD.

In addition, simulations were carried out with various realistic effect sizes for rs2522833 (Figure 2). Compared with the null hypothesis of no replication, the scenario that the non-synonymous SNP, rs2522833, is a causal variant (or in strong LD with it) is 30–50 times more probable, assuming realistic relative risks of 1.2 through to 1.5. An alternative interpretation of these results is that if the true effect of rs2522833 in the Australian cohort were RR = 1.3, the present joint analysis of 29 SNPs has an 87% probability of correctly calling for a true association. Hence, the replication study had sufficient statistical power when performing the joint analysis as proposed here, to confirm the association between PCLO and MDD.

From this reanalysis, we conclude that Sullivan *et al*. provide convincing evidence for the potentially causal association of the non-synonymous coding SNP rs2522833 in PCLO/*Piccolo* with MDD. When combining the evidence from multiple correlated SNPs, we show that the original finding from a GWAS with a nominal *P*-value of 5.4×10^{-7} has been significantly replicated in one cohort. Further studies on PCLO/*Piccolo* in the etiology of mood disorders and a joint analysis of correlated SNPs in finemapping studies are both highly recommended.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

This study was carried out within the framework of the Top Institute Pharma project: number T5-203.

Z Bochdanovits[1], M Verhage[2], AB Smit[3], EJC de Geus[4], D Posthuma[1,4], DI Boomsma[4], BWJH Penninx[5], WJ Hoogendijk[5] and P Heutink[1]
[1]*Department of Medical Genomics, VU University Medical Center, Amsterdam, The Netherlands;* [2]*Department of Functional Genomics, Vrije Universiteit and VU University Medical Center,*

Amsterdam, The Netherlands; [3]*Department of Molecular and Cellular Neurobiology, Vrije Universiteit and VU University Medical Center, Amsterdam, The Netherlands;* [4]*Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands and* [5]*Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands*
E-mail: z.bochdanovits@vumc.nl

References

1 Sullivan PF, de Geus EJC, Willemsen G, James MR, Smit JH, Zandbelt T *et al*. *Mol Psychiatry* 2009; **14**: 359.
2 Terwilliger JD, Hiekkalinna T. *Eur J Hum Genet* 2006; **14**: 426.
3 Bochdanovits Z, Heutink P, van der Vaart A. *Eur J Hum Genet* 2008; **16**: 525.
4 Bacanu S-A, Nelson MR, Ehm MG. *Genet Epidemiol* 2008; **32**: 791.
5 Pardo *et al*. *EJHG* 2009; doi: 10.1038/ejhg.2008.248, advance online publication.
6 Wang T, Jacob H, Ghosh S, Wang X, Zeng Z-B. *Genet Epidemiol* 2008; **33**: 151.
7 Li C, Li M. *Bioinformatics* 2008; **24**: 140.